

| | |
|-----------------|----------------------------------------------------------|
| ชื่อวิทยานิพนธ์ | วิธีการใหม่และปรับปรุงวิธีการสำหรับกราฟข้อมูลอุบัติการณ์ |
| | ประยุกต์ใช้กับข้อมูลการบาดเจ็บจากเหตุการณ์ความไม่สงบ |
| ผู้เขียน | นายสัมพันธ์ คงมาก |
| สาขาวิชา | วิธีวิทยาการวิจัย |
| ปีการศึกษา | 2555 |

บทคัดย่อ

การนำเสนอกราฟในทางสถิติแก่ผู้อ่านที่ไม่ใช่นักสถิติเป็นเรื่องที่ทำหาย โดยเฉพาะผู้อ่านที่มีพื้นฐานทางวิทยาศาสตร์น้อยอาจจะพบว่าการแปลความหมายกราฟในทางสถิติเป็นเรื่องที่ทำความเข้าใจได้ยาก วิทยานิพนธ์ฉบับนี้เป็นการค้นหาวิธีการที่ช่วยผู้อ่านเหล่านั้นให้สามารถเข้าใจความหมายของกราฟทางสถิติได้ง่ายขึ้นโดยการพัฒนาระบบแบบโต้ตอบกับผู้ใช้ขึ้นมา การศึกษาเริ่มโดยการศึกษาการใช้งาน bubble plot, residuals plot, confidence intervals และการฟิตโมเดลทางสถิติซึ่งเป็นพื้นฐานในการวิเคราะห์ข้อมูล

วิทยานิพนธ์เล่มนี้ประกอบด้วยบทความที่ได้รับการเผยแพร่แล้ว จำนวน 2 ฉบับ โดยที่ฉบับแรกเป็นการเปรียบเทียบการฟิตโมเดลทางสถิติกับข้อมูลอัตราอุบัติการณ์ของข้อมูลการบาดเจ็บจากสถานการณ์ความไม่สงบในพื้นที่สามจังหวัดภาคใต้และ 4 อำเภอของจังหวัดสงขลา โดยมุ่งเน้นการศึกษาข้อมูลผู้ที่ไม่ได้นับถือศาสนาอิสลาม (ส่วนใหญ่นับถือศาสนาพุทธ) เนื่องจากประชาชนกลุ่มนี้ดูเหมือนว่าจะเป็นเป้าหมายของการก่อการร้ายมากกว่าผู้นับถือศาสนาอิสลาม ด้วยความอนุเคราะห์ข้อมูลจากศูนย์ประสานงานวิชาการให้ความช่วยเหลือผู้ได้รับผลกระทบจากเหตุการณ์ความไม่สงบจังหวัดชายแดนภาคใต้ (ศวชต.) ข้อมูลตั้งแต่ปี พ.ศ. 2547 ถึง พ.ศ. 2553 โดยนับอัตราอุบัติการณ์ของผู้เคราะห์ร้ายต่อประชากร 100,000 คน โดยสนใจอัตราอุบัติการณ์ที่เกิดขึ้นต่อช่วงเวลา ลักษณะเฉพาะของประชากรศาสตร์ เพศและกลุ่มอายุซึ่งแบ่งออกเป็น 3 กลุ่ม ประกอบด้วย กลุ่มที่อายุน้อยกว่า 25 ปี กลุ่มอายุระหว่าง 25 ถึง 44 ปี และกลุ่มที่มีอายุตั้งแต่ 45 ปีขึ้นไป กลุ่มอายุน้อยกว่า 25 ปีคือกลุ่มที่อยู่ในวัยเรียน กลุ่มที่มีอายุระหว่าง 25 ถึง 44 ปีคือกลุ่มวัยทำงาน และกลุ่มที่มีอายุตั้งแต่ 45 ปีขึ้นไปคือกลุ่มที่อยู่ในวัยผู้บริหารซึ่งเป็นกลุ่มที่มักตกเป็นเป้าหมายของ

การก่อการร้าย ข้อมูลทั้งหมด 37 อำเภอ ขึ้นตอนเริ่มด้วยการรวมอำเภอติดกันเป็น 23 เขต แต่ละเขตจึงมีจำนวนประชากรใกล้เคียงกัน อย่างไรก็ตามเมื่อพิจารณาต่อไปก็พบว่าแต่ละเขตยังมีจำนวนผู้ที่ไม่นับถืออิสลามแตกต่างกันมาก จึงรวมใหม่อีกครั้งจนเหลือ 15 เขต จากนั้นทดลองฟิตโมเดลด้วยโมเดลต่อไปนี้คือ Poisson regression, negative binomial regression และ linear regression ซึ่งเรียกว่า log-normal ผลจากการวิเคราะห์พบว่า Poisson regression ได้ผลดีที่สุด ในขณะที่อีกสองโมเดลได้ผลดีกว่า โดยเฉพาะ log-normal ได้ผลดีที่สุด ดังนั้น log-normal จึงถือว่าเป็นโมเดลที่เหมาะสมสำหรับประมาณการณ์ข้อมูลผู้รับเคราะห์จากการก่อการร้ายมากที่สุด

บทความที่สองได้มีการพัฒนาวิธีการใหม่สำหรับการแสดงผลข้อมูลและสร้างกราฟสำหรับการวิเคราะห์ทางสถิติ แบบสามมิติ ด้วย Google Earth และภาษา KML (Keyhole Markup Language) ที่สามารถใช้งานได้ฟรีจากอินเทอร์เน็ต เป้าหมายหลักคือการสร้างวิธีการแสดงผลผลลัพธ์ทางสถิติที่ซับซ้อนให้ผู้อ่านสามารถเข้าใจความซับซ้อนของข้อมูลได้ง่ายขึ้น การศึกษาเริ่มด้วยการเลือกข้อมูลจาก สวชต. จากปีพ.ศ. 2547 ถึง พ.ศ. 2552 ที่เกี่ยวกับเหตุการณ์ความไม่สงบในสามจังหวัดชายแดนภาคใต้และ 4 อำเภอของจังหวัดสงขลา เนื่องจากจำนวนอุบัติเหตุที่เกิดขึ้นกับผู้นับถือศาสนาอิสลามมีจำนวนน้อยกว่ามาก เมื่อเทียบกับจำนวนผู้นับถือศาสนาอื่น เลยเป็นเหตุให้นเน้นการศึกษาอัตราอุบัติเหตุของผู้นับถือศาสนาอื่น การศึกษาข้อมูลอุบัติเหตุการก่อการร้ายแบ่งออกเป็นปัจจัยต่อไปนี้คือ เพศ กลุ่มอายุ (น้อยกว่า 25 ปี 25 ถึง 44 ปี และ 45 ปีขึ้นไป) และพื้นที่ที่เกิดเหตุ วิธีการเริ่มจากรวมกลุ่มอำเภอที่อยู่ใกล้ ๆ กัน จาก 37 อำเภอ ไปเป็น 23 เขต แต่ละเขตจะมีประชากรระหว่าง 54,039 ถึง 154,634 แต่ตอนหลังพบว่าบางเขตมีจำนวนประชากรที่นับถือศาสนาอื่นที่ไม่ใช่อิสลาม น้อยกว่า 5,000 คนต่อเขต จึงได้มีการรวมเขตใหม่อีกครั้งจนเหลือ 15 เขต จึงทำให้สามารถประมาณจำนวนอุบัติเหตุที่น่าเชื่อถือทางสถิติได้ จากนั้นทดสอบแฟลคเตอร์ทั้ง 4 ที่กล่าวมาแล้วนั้นด้วยโมเดลการถดถอย ด้วยการลอการิธึมเพื่อลดความเบ้ของข้อมูล เพื่อให้มีความน่าเชื่อถือในทางสถิติ อย่างไรก็ตามมีข้อมูลจำนวน 109 ช่องจาก 540 ช่องที่เป็น 0 ดังนั้นจำเป็นต้องแทนที่ 0 ด้วย 0.5 เพื่อให้สามารถทำการลอการิธึมได้ และผลจากการวิเคราะห์ในรูปแบบ confidence intervals พบว่าผู้ชายมีความเสี่ยงมากกว่าผู้หญิง แต่เนื่องจากการอ่านผลจาก confidence intervals นั้นยากต่อการเข้าใจโดยคนทั่วไป เราจึงสร้าง Thematic Map เพื่อให้คนทั่วไปสามารถเข้าใจได้ โดยให้สีแดงคือพื้นที่อันตรายมาก สีเขียวคืออันตรายและสีน้ำเงินคือพื้นที่อันตรายน้อย

หลังจากนั้นก็สร้างระบบใหม่ชื่อว่า “Interactive GIS graphics System” เพื่อนำเสนอข้อมูลในลักษณะกราฟแท่งด้วยสีต่างๆ เหมือน Thematic Map และเพิ่มความสามารถในการตอบโต้กับผู้ใช้เข้าไปด้วย รวมทั้งกำหนดสีพื้นที่ให้กับแผนที่เพื่อบอกระดับความอันตรายด้วย Google Earth ด้วยภาษา KML ภาษา KML เป็นภาษาที่มีโครงสร้างคล้ายกับภาษา HTML แต่ปัญหาคือไม่สามารถเปลี่ยนแปลงข้อมูลในแฟ้มภาษา KML ได้ ดังนั้นจึงแก้ปัญหาโดยการเขียนโปรแกรมด้วยโปรแกรม R เพื่อให้เปลี่ยนแปลงข้อมูลในแฟ้มภาษา KML โดยอัตโนมัติ ในที่สุดจึงได้ระบบที่สามารถตอบโต้กับผู้ใช้ รวมถึงความสามารถในการซูม การเปรียบเทียบ หรือการมองภาพมุมสูง

Prince of Songkla University
Pattani Campus

| | |
|----------------------|-------------------------------------------------------------------------------------------------|
| Thesis Title | New and Improved Methods for Graphing Incidence Data, with Application to Terrorism Injuries |
| Author | Mr. Sumpunt Khongmark |
| Major Program | Research Methodology |
| Academic Year | 2012 |

ABSTRACT

Presenting statistical graphs to the viewer who is not a statistician is challenging, especially a viewer who has insufficient science knowledge, and they might find the graphs difficult to understand. This thesis will show a solution to assist them to understand the meaning of such graphs by developing interactive systems. The study started by using statistical graphs such as the bubble plot, residuals plot, confidence interval plot, and fitted statistics models, as in conventional statistical analysis.

The thesis contains two published articles summarizing the main results. The first article compares statistical models of incidence rates of terrorism injuries in the three southern-most provinces of Thailand and four districts from Songkhla province. The target of terrorism data is focused on non-Muslim (mostly Buddhist) because they seem to be targeted more than Muslim based on data from the Deep South Coordination Center (DSCC), for years 2004 to 2010, using incidence rates per 100,000 population for civilian victims of terrorism. We compared incidence rates of events that happen over space, time, and demographic characteristics of subject including gender and age group classified into 3 groups consisting of less than 25 years, 25 to 44 years, and 45 years of age and more. We classified like this because those less than 25 years of age

usually are in education, 25-44 year olds are working, and 45 years of age and more are mostly senior people who seem to be more targeted than others. The data that we are interested in are for 37 districts. Firstly we grouped them into 23 regions by joining some small districts together, making all regions approximately the same population size. However, the population of non-Muslims were still different for each region then we grouped them again ending up with 15 regions. Furthermore, we considered three different statistical methods: the Poisson regression, negative binomial and linear regression with log-transformed, called log-normal. The results show the Poisson model fitted poorly and the others are better. However, the log-normal model seems to fit the best for this data set. That means the log-normal model is the best for estimating the proportion of terrorism victims.

In the second article we developed a new method to visualize data and statistics results over three-dimensional space using Google Earth and coding in KML (Keyhole Markup Language) which was created dynamically using R program (Google Earth and KML are freely available to download and use from the Internet). The main target of this method is to provide a more effective method for displaying complex statistical results, such as those derived in the first article, in a way that enables the viewer to more easily and quickly digest such complex information. Firstly, the article started with selected data from the Deep South Coordination Center (DSCC) from 2004 to 2009 about terrorism in the three southern-most provinces of Thailand and four districts from Songkhla province. Because the overall rate of victim incidence for Muslims seems much less than non-Muslim we restricted the study to the non-Muslim population. The terrorism events are again classified by gender, age group (<25, 25-44, or 45 years old more), and area of incidence. We first aggregated 37 districts into 23 regions by

merging small and close districts together. Each region contains population ranging from 54,039 to 154,634. Finally, we found, non-Muslim residence from some region has less than 5,000. So, at the end, we again aggregated into 15 regions to provide acceptably stable estimates of incidence rates. Secondly, we used four factors above to fit with regression model, taking log-transformation to remove skewedness of data and thus satisfy statistical assumptions. However, there were 109 of 540 cells containing zeroes, which we replaced by 0.5 to enable log-transformation. The result shows females have lower risk than males. However, showing the confidence interval plot is difficult to be understood by people who have no statistics background. Then we made a thematic map with different colors for locations using information from the confidence intervals. Red, blue, or green colours were used. Then we replaced the thematic map with interactive coloured bar charts as “Interactive GIS Graphics System”, group three-dimensional bar charts. The idea is to put a coloured bar chart over the real location on Google Earth by using KML. KML is a static tag-based language similar to Hyper Text Markup Language (HTML) which extends from XML. The problem of KML is that the KML code needs to be changed whenever the data are changed. We solved the problem by employing R programs to generate KML, dynamically creating a KML file containing coloured three-dimensional bar charts, and screen overlay confidence intervals. The resulting KML file can be opened with Google Earth. Finally, we have a system that can be interactive, including zooming, turn around, comparison, and even making a bird eye view.